ELSEVIER

Agricultural Economics 27 (2002) 295–315

AGRICULTURAL ECONOMICS

www.elsevier.com/locate/agecon

Location and uptake: integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya

S.J. Staal [a,*], I. Baltenweck [a], M.M. Waithaka [b], T. deWolff [a], L. Njoroge [a]

[a] *International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya*
[b] *Kenya Agricultural Research Institute, P.O. Box 57811, Nairobi, Kenya*

Abstract

GIS-derived measures of location and space have increasingly been used in models of land use and ecology. However, they have made few inroads into the literature on technology adoption in developing countries, which continues to rely mainly on survey-derived information. Location, with all its dimensions of market access, demographics and agro-climate, nevertheless remains key to understanding potential for technology use. The measures of location typically used in the adoption literature, such as locational dummy variables that proxy a range of locational factors, now appear relatively crude given the increased availability of more explicit GIS-derived measures. This paper attempts to demonstrate the usefulness of integrating GIS-measures into analysis of technology uptake, for better differentiating and understanding locational effects. A set of GIS-derived measures of market access and agro-climate are included in a standard household model of technology uptake, applied to smallholder dairy farms in Kenya, using a sample of 3330 geo-referenced farm households. The three technologies examined are keeping of dairy cattle, planting of specialised fodder, and use of concentrate feed. Logit estimations are conducted that significantly differentiate effects of individual household characteristics from those related to location. The predicted values of the locational variables are then used to make spatial predictions of technology potential. Comparisons are made with estimations based only on survey data, which demonstrate that while overall explanatory power may not improve with GIS-derived variables, the latter yield more practical interpretations, which is further demonstrated through predictions of technology uptake change with a shift in infrastructure policy. Although requiring large geo-referenced data sets and high resolution GIS layers, the methodology demonstrates the potential to better unravel the multiple effects of location on farmer decisions on technology and land use.

Keywords: Spatial analysis; Geo-referenced farms; Technology adoption; Market access; Geographical information system

1. Introduction

The often-heard refrain as to what determines the value of a piece of residential or commercial land is “location, location, location” (Geoghegan et al., 1997). It follows then, that the technologies that are employed in association with that land as well, should be equally affected by its location in all its manifestations. It can be argued however, that the rich field of agricultural economics literature devoted to understanding the uptake of technology has not adequately incorporated locational effects in its analyses. This in spite of the fact that since von Thünen’s time (Von Thünen, 1966), it has been clearly recognised that

* Corresponding author. Tel.: +254-2-630743; fax: +254-2-631499.
E-mail address: s.staal@cgiar.org (S.J. Staal).

PII: S0169-5150(02)00075-0

most economic activity has a strong spatial component (1966). The tools typically employed to capture locational and spatial effects in adoption analysis are dummy variables to differentiate regions, or approximate distances to urban or market centres. Dummy variables may capture a wide range of locational effects, from soil to climate to infrastructure, which can be impossible to differentiate and difficult to interpret. GIS tools provide new measures of location and spatially-differentiated variables that may be able to more explicitly quantify effects of spatial factors on technology uptake and land use. Further, while GIS-based land-use analysis has emphasised environmental and resource issues, more conventional issues of economic behaviour that also have spatial components may have been overlooked to date. It is the aim of this paper to demonstrate that GIS-derived measures of spatially-differentiated factors can be incorporated into the standard household adoption model, and can potentially differentiate the multiple aspects of location on choices of agricultural technology. The method is applied to a key farming system for east Africa, smallholder dairy farming. Dairy systems are particularly suitable to spatial analysis, given their heavy reliance on markets for a perishable product, and the role that agro-climate can play in driving productivity.

2. Location

As established in the literature, technology uptake is driven conceptually by a desire to maximise farmer utility in the context of individual and household resources, incentives presented by the external environment, and perceptions of the technology and of the risks associated with it (Feder et al., 1985; Kaliba et al., 1997; Sall et al., 2000). Location can play a role in several of these categories of determining factors, affecting either demand or supply of technology. Agro-climatic effects such as level of rainfall and its distribution, or soil and temperature, condition the household resources and determine the basic agricultural productivity of the land. Other spatial factors, such as market access, condition the incentives presented by the external environment at a certain location, whether through inputs, outputs, or other services. The demographic space is that related to neighbours and their land uses, the patterns and densities thereof, and resulting externalities. This concept has been employed in spatial studies of residential property value that measure neighbouring open space and its fragmentation (Geoghegan et al., 1997). In the context of developing country technology, neighbouring spaces and populations may affect transmission of knowledge, and access to common property. Even more abstract surfaces, related to the dynamic process of diffusion of information and technology, may be institutional structure surfaces (Bockstael, 1996). This may refer to social institutions that condition preferences and change gradually across ethnic or community boundaries, or perhaps organisational structures or policies that change abruptly with administrative boundaries.

3. Market access and its measurement

A focus of this paper is on market access and its measurement, largely because market-oriented dairy systems have traditionally been heavily dependent on infrastructure, and on spatial thresholds of milk surplus and deficit. Additionally, it is our contention that the environmental and land-use literature which most employs GIS tools has emphasised agro-climatic surfaces, while incorporating market access measures at a poorer, more coarse level of resolution. As we hope to demonstrate, high-resolution market access measures may yield significant inferences for the behaviour of farmers whose main means of transport may be a bicycle.

Generally, the concept of market access combines several elements: (1) distance between the point of observation and some market destination or combinations of destinations, (2) the utility of the market destinations, based on their supply or demand attributes, and (3) the impedance level or quality of the route, in terms of relative ease of movement for goods, services, people and perhaps even information. Combined, at the point of observation, these three elements condition prices observed at that location for inputs, outputs and services, their availability, and even their quality. Indeed, when prices cannot be observed, market access measures can serve as a proxy for price (Chomitz and Gray, 1996). Finally, in keeping with transaction cost theory (Williamson, 1985), it is useful to add another element to market access, (4) the individual capacity and resources of the agents at the point of observation. Differences in transactions costs faced at any point are partially a function of the individual characteristics of the market agents involved, including ability to use market information and to conduct exchanges, thus partially determining market access at that point.

In measuring market access, some or all of these elements can be used. The simplest measures may use only distances to destinations, or may differentiate those distances by level of impedance. A typical approach is to assume or estimate typical travel speeds for specific types of routes and surfaces, or to use their inverse ratios (Chomitz and Gray, 1996; Nelson and Hellerstein, 1997). In a GIS the access measures can be calculated using road networks connecting points with specific destinations yielding continuous measurement of access, or by creating buffers around points of a specific distance, creating categories of access (Serneels and Lambin, 2001). When the utility of market destination is added to these weighted distance, composite indices of market access are derived. These range from the classical gravity model of market access, to more flexible forms such as the negative exponential model (Deichmann, 1997), which allows flexibility in rate of decay. The composite potential measures may reflect a more complex interaction of market factors, but are unit less, and difficult to interpret. de Wolff et al. (2000) examine and test these various measures, and find that simple distance measures may be more useful in measuring market access, depending on the type of market and commodity being addressed. As will be described, in this study we employ combinations of relatively simple measures of distance to urban areas, differentiated by road type.

4. Measures of location in the literature

Given the multiple dimensions of location and market access, the measures of location used in traditional economic studies may be regarded as relatively crude proxies. They typically consist of dummy variables for location, represented by the administrative unit such as village or county, and simple farmer or researcher estimates of distance to an urban centre or road. Such measures have shortcomings, in that they proxy a variety of spatial factors ranging from market and institutional access to agro-climate, cultural and historical variation. The interpretation of the results obtained requires a fair bit of speculation as to which of these factors are associated with the observed outcomes. While some studies employ such locational dummy variables (Kaliba et al., 1997), others use a regional measure of productivity (Feder and Slade, 1984), or conduct separate estimates for different locations (Lapar and Pandey, 1999). At the end of the scale, many farm technology uptake studies include no explicit spatial variables at all (Rahm and Huffman, 1984; Polson and Spencer, 1991; Nkonya et al., 1997; Adesina et al., 2000; Sall et al., 2000). This may be because the economic agents are seen as discrete distinct decision-makers, but are not recognised to exist in an explicit spatial context (Anselin, 2001).

In contrast to the non-spatial approach usually taken by economic studies, environmental and land-use studies have explicitly treated the spatial dimensions, clearly because the focus on land requires attention to the wider landscape and to location. Many have also linked GIS-derived spatial information with socio-economic variables, typically using spatial grid data (de Koning, 1999; Verberg and Chen, 2000; Reid et al., 2000) or *field* data as described by Anselin in this issue. Two land-use studies in particular, Chomitz and Gray (1996) and Nelson and Hellerstein (1997), provide a method that can be adapted to the analysis of technology adoption. Both use an economic framework, and multinomial logit regression models similar to those used in household adoption analysis, but now applied to spatial grid data obtained by remote sensing along with GIS-derived market access measures. Nelson and Hellerstein (1997) also underline the importance of measuring route quality, by demonstrating that impedance-weighted road distances provide generally better market access measures than simple vector distances.

5. Integrated household and GIS based analysis of technology adoption

The above models of land-use change are aimed at identifying where or at what rate land-use change is taking place. From the technology adoption perspective, if we assume exogeniety of land resources and location to individual households, the question is *what* rather than *where*. Very similar approaches can be used for both sets of questions. The key is to more effectively integrate spatially-differentiated measures of the non-physical social and economic landscapes with the physical, regardless of whether addressing physical or behavioural outcomes.

The approach employed in this paper for dealing with this problem is to integrate spatially referenced household data (*objects*, Anselin in this issue) with point data derived from digital surfaces and infrastructure maps (*field* data). A key difference from the approach used in the above studies is the unit of observation, which is a household, rather than a spatial grid cell or administrative unit. As has been pointed out by others, a key to linking household and GIS data is to correctly define the spatial observation unit (Mertens et al., 2000; Geoghegan et al., 2001). Administrative units or grid cells are not individual economic agents, but simply aggregates of them (Anselin, 2001). Inferences as to outcomes in such units require simplifying assumptions about homogeneity of the decision-makers and the dynamics comprising that aggregate, a constraint that plagues land-use analysis in a systematic manner. Anselin (2001) indicates that proper inference of micro-level spatial behaviour is therefore more appropriately based on survey samples of individual agents, under the general principal of matching the spatial scale of the process under consideration and the scale at which measurement is carried out. The approach applied here does that, linking spatial measures to the perceived real decision-makers, thus matching the spatial and behavioural units. The related scale issue centres around the problem of optimally defining the spatial entities or regions for which a model should be calibrated, depending on the level of heterogeneity within units chosen: the greater the variability, or the smaller the spatial scale at which the process operates, the less accurate will be the aggregate as an estimate for the dependent variable (Anselin, 2001). For the small farmers with limited access to motorised transport under study here, we thus apply a high resolution road network.

While analysis of residential areas increasingly integrates GIS data (Geoghegan et al., 1997; Irwin and Bockstael, 2001), few studies have integrated household and GIS point data in this manner applied to agricultural development, such as Mertens et al. (2000) and Swallow et al. (2000). Mertens et al. use household-derived data linked to remotely sensed data to examine the impact of macroeconomic changes on deforestation in Cameroon. They use a village level of analysis and so aggregate household and plot data to that level. In another livestock related study, Swallow et al. use GIS tools applied to household survey data to examine livestock disease control technology uptake in Ethiopia. They rely on GIS-derived "neighbourhood" variables to explain access to information and markets, although they also employ simple vector distances and locational dummy variables. At a minimum, GIS data make distance and other spatial measures more accurate or more easily computable (Anselin, 2001). More significantly though, integrating GIS data into household adoption analysis also offers the possibility of better differentiating the multiple effects of location, and also of predicting potential outcomes under different policy and infrastructure scenarios. This analysis attempts to demonstrate that in the case of several technologies employed by smallholder dairy farmers in Kenya.

6. Smallholder dairy production in Kenya

In Kenya smallholder dairy farmers produce some 56% of total milk and 80% of the total marketed milk nationally (Omore et al., 1999). In the study area, which comprises the main milk producing regions of the country (Fig. 1), about 87% of all households were farms, and of these 72% had dairy cattle. Most of the districts surveyed ranked dairy as the main source of farm income (Staal et al., 2001; Waithaka et al., 2000). Smallholder dairying has thrived on the good agro-climatic conditions in the temperate Kenyan highlands, most of which are located over 1500 m above m.s.l. Its success has also been due to several decades of government support through a dairy parastatal that until recently, acted as guaranteed buyer of milk output. In the study area, dairy production is typically conducted on a few acres, with a herd of crossbred cows ranging from 1 to 5 in size. Production is based on the close integration of dairy cattle into the mainly maize-based farms, and is sometimes accompanied by cash crops such as coffee, tea, or pyrethrum. The cattle are usually Friesian or Ayrshire crossed with local Zebu. An important element of this system is the use of the manure to fertilise food and cash crops, allowing sustained multiple cropping on the

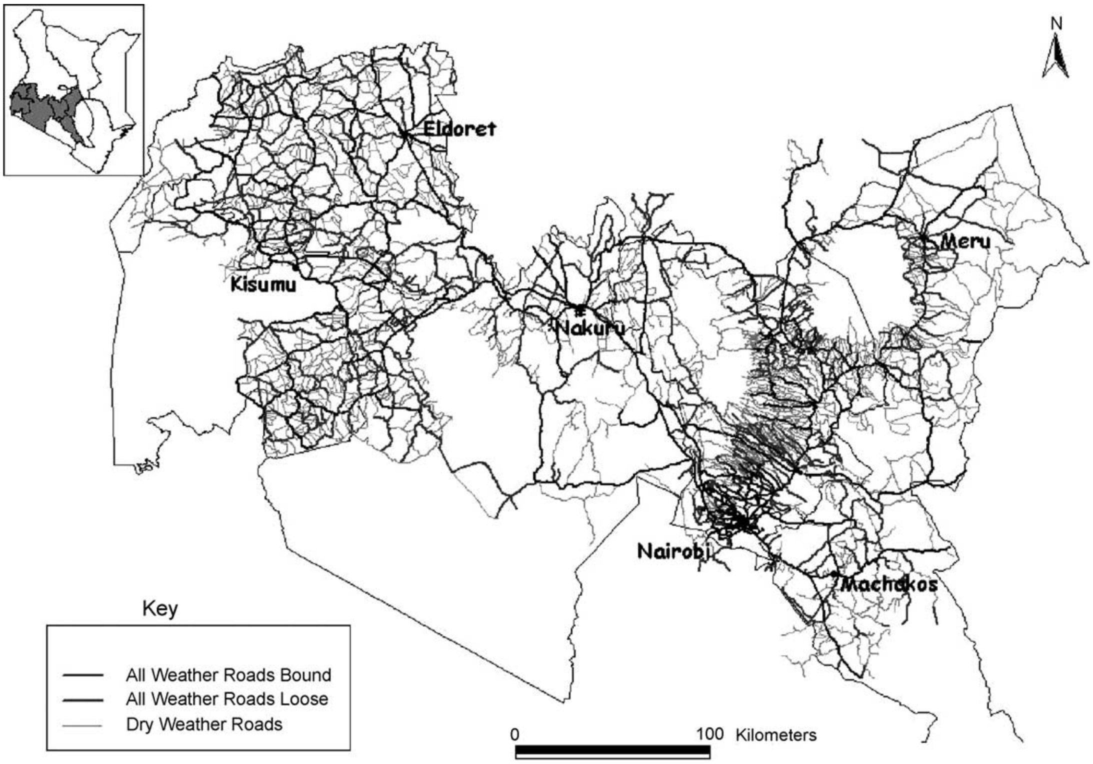

Fig. 1. Household survey area and road network map.

smallholdings. Cattle are fed planted fodder (*Pennisetum purpureum*, Napier grass), maize stover, weeds and grass, and sometimes supplemented with concentrate feeds such as grain millings or compounded dairy feeds. It is important to note that in some cases, a large proportion of fodder is gathered from public or common land or is purchased, so feed resources are by no means limited to those produced on farm. The main feeding systems in the area are stall-feeding based on cut- and carry of fodder in the case of some 32% of dairy farms, and only grazing for 39%, with the remainder employing some combination of the two. Milk production per animal is low, typically 4–7 l per day. The main cattle disease challenge is a tick-borne disease, east coast fever (ECF), which causes mortality particularly in herds that graze, whose exposure to ticks is greater.

Key to understanding milk production in Kenya is recognition of the role of the informal, or raw milk, market. It is estimated that about 80% of marketed milk is neither processed nor packaged, but is bought by the consumer in raw form (Omore et al., 1999), mainly due to traditional preferences for fresh raw milk, and to the unwillingness of resource-poor consumers to pay the costs of processing and packaging. Thus, the largest single market outlet for farmers, or 36% of marketed milk, consists of direct sales to nearby households, followed by small traders, who deliver 28% of the marketed milk to consumers or other retail outlets. Private formal dairy processors capture only some 19% (Omore et al., 1999).

On the input side, dairy farms are dependent on livestock services such as clinical veterinary services and artificial insemination (AI), critical to maintaining the health of susceptible cross-bred cattle and the genetic potential for higher milk yields. Since the early 1990s, however, the government has reduced support to AI and veterinary services. The expected replacement of these public services by private entrepreneurs has only partially occurred, and then only in high cattle density areas where demand is adequate to support them.

The important locational characteristics to smallholder dairy farming thus have to do with agro-climate, demographics, and access to both output markets and those for livestock services and feeds. Average levels of rainfall and temperature largely determine the feed production potential, not only of a farmer's own land, but of the public lands from which fodder is often gathered. The density of neighbouring households will partially determine the extent to which public lands are available for fodder gathering. For farmers with limited means of transport and communication, distance to livestock service centres will affect timeliness and thus effectiveness of such services, and the costs of accessing them. Finally, for a highly perishable product such as milk, failure to reach the market point in time may mean the loss of that day's revenues. The transfer and transaction costs associated with milk sales will generally be related to distance to consumer markets, distance to milk collecting and selling points, and the road infrastructure. Since many milk sales are to neighbours, local net demand and potentially, population density, will also be a determinant.

7. Data sources

7.1. Household surveys

Household data were obtained from three surveys conducted in central and western Kenya between 1996 and 2000, as part of an effort to characterise smallholder dairy systems by a collaborative team from the Ministry of Agriculture and Rural Development, the Kenya Agricultural Research Institute (KARI), and the International Livestock Research Institute (ILRI). Similar sampling methods were applied in each case, and each survey used a variant of the same data collection instrument, conducted in a single interview of each household. The survey collected a wide variety of data on household resources, land use and livestock management practices, livestock inventory, recall of feed and other input use, and use of livestock and extension services.

Prospective study districts were grouped according to agro-ecological production potential (high or medium) and market access (high, medium, or low). One to two districts were chosen within each combination, within which density classes were identified, from which two sub-locations, the smallest administrative unit, were randomly selected. Sample sizes were weighted by household estimates extrapolated from the 1989 census figures, and a sample size totalling 3330 was obtained. Random transects were then drawn in each sub-location, and every fifth household along the transect was selected until the desired samples were obtained, whether a farm household or not. Each household was geo-referenced using a GPS unit. All main milk processing and collecting centres in the study area were also geo-referenced. Mean values for the variables used in the analyses are shown in Table 1.

7.2. GIS layers

The primary new GIS coverage needed for the analysis was a detailed road network of the area, for which digitised maps at the level of resolution required were not available. Topographic map sheets at a scale of 1:50,000 were acquired from the Survey of Kenya to cover the study area and three classes of roads were digitised: (1) all-weather, paved surface (tarmac), (2) all-weather, loose surface, and (3) dry weather only (Table 2). Information was obtained from district-level road authorities on recent road renovation, and all main roads were visited to update the quality attributes in the GIS. GIS software (workstation ARC/INFO, ESRI, 1998) was then utilised to assign farm or facility information to the nearest node or intersection in the network, and major urban areas such as Nairobi and other towns were added as nodes, as were the milk market facilities. The resulting network contains a total of 10,199 nodes and 11,488 road sections. The road sections in the network were then assigned a quality variable that reflects assumed mean travel speeds, with values ranging from 30 km/h in the case of dry weather only roads (type 3) to 80 km/h for hard surface roads (type 1) (Table 2). The GIS was then used to calculate travel times based on each section's length and its associated travel speed, which then were used to identify least travel-time routes. To do this, the Arc/Info GIS-network module (ESRI, 1998) was used, and for each node on the network were obtained: (a) total distance to the largest city (Nairobi) by least travel-time path, (b) distance to the two nearest urban areas by least travel-time paths, separated by road type, and (c) distance to the nearest formal milk collection centres

Table 1
Description of independent variables included in household technology adoption model, source, expected sign and rationale for inclusion

Variable	Data source	Ho sign	Rationale
Sex of the household (hh) head	Survey	+	Male headed households are likely to have better access to information and services
Years of farming experience for the hh head	Survey	+	Longer farming experience predisposes farmer to better farming techniques through learning by doing
Years of formal education, hh head	Survey	+	More formal education is likely to increase farmer capacity for management and for utilising information
No. of adults in the hh	Survey	+	More adults implies more farm labour for the labour intensive dairy activity
Ratio of adult female to total adults in hh	Survey	+	Women contribute proportionately more labour to dairy and fodder production, and therefore a proportionately high number of women implies higher adoption
Ratio of hh members below 14 and above 65 to total family size	Survey	Uncertain	A higher proportion of dependents in the household implies less labour for dairy and fodder production. On the other hand, a higher dependency ratio gives incentives to produce more milk for consumption
Acreage under maize	Survey	(+) for dairy cows, (−) for Napier and concentrate	Higher maize acreage leads to greater fodder availability from crop residues
Hh landholding size in acres	Survey	(+) for dairy and Napier, (−) for concentrate	Large land size may mean there is more land available for cattle keeping and growing fodder, but may reduce need for concentrate feeding
Percentage of hh in the sub-location with stated access to animal health services	Survey	+	Availability of animal health provides support for the dairy enterprise in general, and all dairy-related technologies
Percentage of hh in the sub-location with stated access to formal milk outlets	Survey	+	Local availability of formal milk outlets implies reliable milk market, and support for dairy enterprise
Percentage of hh in the sub-location with stated access to extension services	Survey	+	Availability of extension services implies support for the dairy enterprise in general
Distance to the nearest formal milk collection centre by type 1 road (tarmac)	GIS	−	Distance to milk collection centre is expected to reduce formal milk price and market reliability
Total distance by road from homestead to Nairobi (km)	GIS	−	Nairobi is the largest milk market, particularly for the formal market. Greater distance to that market is expected to reduce price and market reliability
Mean human population density, 5 km radius	GIS	+	High population density is expected to correspond to higher local demand for milk, and thus favourable local market
Annual precipitation/potential evapo- transpiration (PPE)	GIS	+	Higher PPE corresponds to more favourable agro-climatic conditions for dairy production
Tarmac road distance along routes to two the nearest urban centres	GIS	−	Greater distance along main roads to nearby urban centres is expected to reduce prices in milk markets, and raise prices and reduce availability of inputs and services
All-weather earth road distance along routes to the nearest urban centres	GIS	−	Greater distance along secondary roads is expected to reduce milk prices and raise prices and reduce availability for inputs and services
Dry-weather only road distance along routes to two the nearest urban centres	GIS	−	Greater distance along seasonal roads to main roads will reduce access to input and output markets, and increase seasonal risks

Table 2
Description of household survey and GIS-derived variables included in the household models

Description	Cattle keeping households ($n = 2048$)		Agricultural households ($n = 2864$)	
	Mean	S.E.	Mean	S.E.
Dependent variables				
Cattle keeping: 1 if hh has any cross or grade cattle, 0 otherwise			0.62	0.49
Napier cultivation: 1 if Napier is planted, 0 otherwise	0.53	0.50		
Concentrate use: 1 if any concentrate is used, 0 otherwise	0.25	0.43		
Concentrate use intensity: Total kg of concentrates per year/total no. of cross + grade cows	157.63	491.95		
Household head characteristics (household survey)				
Sex of the hh head: 1 if male, 0 if female	0.79	0.41	0.78	0.41
Years of farming experience of the hh head	21.23	13.29	19.71	13.52
Years of education of the hh head	7.93	4.55	7.88	4.52
Household characteristics (household survey & GIS)				
No. of adults in the hh	3.84	2.11	3.59	2.03
Ratio of adult females to total adults in hh	0.53	0.21	0.53	0.21
Ratio of hh members below 14 and above 65 to total family size	0.41	0.25	0.42	0.25
Acreage under maize	0.87	1.13	0.78	1.03
Land size in acres	5.45	9.34	4.53	8.11
Annual precipitation/potential evapo-transpiration (PPE)	0.91	0.22	0.91	0.22
Neighbourhood characteristics (household survey)				
Percentage of hh in the sub-location with stated access to animal health services	92.31	11.83	92.02	12.57
Percentage of hh in the sub-location with stated access to formal milk outlets	16.86	28.98	14.86	27.57
Percentage of hh in the sub-location with stated access to extension services	90.13	18.69	89.66	19.54
Market infrastructure (GIS)				
Distance to the nearest formal milk collection centre by type 1 road (tarmac)	15.16	18.74	16.25	18.88
Total distance by road from homestead to Nairobi (km)	225.74	135.05	233.66	135.19
Mean human population density, 5 km radius	480.38	311.61	487.90	330.28
Tarmac road distance along routes to two the nearest urban centres	22.17	13.41	21.70	13.30
All-weather earth road distance along routes to the nearest urban centres	6.52	7.06	6.47	6.91
Dry-weather only road distance along routes to two the nearest urban centres	2.63	3.40	2.51	3.32

by least travel-time paths, by road type. Interpolation was used to produce smoothed accessibility surfaces for the whole study area. Arcview 3.1 Spatial Analyst (ESRI, 1999) was used to accomplish these interpolations, which utilises a simple inverse weighted distance algorithm.

The main agro-climatic variable used was precipitation/potential evapo-transpiration (PPE), which is an index that combines average effects of rainfall, altitude, and sun radiation, obtained from the Almanac Characterisation Tool (Corbett, 1999). The human population density layer was developed at ILRI and is based on the 1989 Kenya census, and is attached to sub-location boundaries. Using Arcview Spatial Analyst, focal neighbourhood functions were used to evaluate the mean population density within a 5 km radius for every point in the study area. Mean values for all of the above variables are shown in Table 2.

8. The integrated household model

8.1. Theoretical model

This analysis uses the same broad approach as used by many in this literature set; our emphasis here is on the empirical side, in introducing a new set of measures for spatial variables. Following Besley and Case (1993), the farmer i adopts the new technology if the derived benefits B_i are higher than a certain threshold

T. The decision to adopt can then be written as:

$$Y_i = 1 \text{ if } B_i > T \Rightarrow X_i\beta + \varepsilon_i > T$$
farmer i decides to adopt (1)

$$Y_i = 0 \text{ if } B_i < T \Rightarrow X_i\beta + \varepsilon_i < T$$
farmer i decides not to adopt (2)

where X_i is a vector of explanatory variables, β a vector of coefficients to be estimated and ε_i is an independently and identically distributed farm specific ex ante shocks. In the case of adoption of agricultural innovations in developing countries, Feder et al. (1985) identify key characteristics to be included in X_i farmer's characteristics, farm characteristics and external factors. Our addition to the model is simply to empirically express some farm characteristics and external factors using GIS-derived variables.

8.2. Empirical model

The model estimated is thus of the form:

$$Y_i = x_i\beta_1 + z_i\beta_2 + \varepsilon_i \quad (3)$$

where x_i is a vector of explanatory variables derived from household surveys, with β_1 as the corresponding regression coefficients, and z_i is another vector of explanatory variables derived from GIS surfaces, and β_2 the corresponding coefficients. Both types of variables are evaluated at the household level.

As Anselin (2001) points out, however, the values of z_i are not actual observations, but are instead predicted values generated through the spatial interpolation described earlier. As such, z_i may have its own error structure, additional to ε_i, and potentially not independent from it. Correlation with ε_i would be more likely if there are similar spatial patterns in the two error terms. Anselin suggests that this should be addressed by means of instrumental variables, as employed by Chomitz and Gray (1996). In our case, no adequate variables were seen to be available to instrument the main spatial factors, weighted distances and agro-climate.

8.3. Spatial autocorrelation

Spatial autocorrelation refers to the "lack of independence which is often present among observations in cross-sectional data sets" (Anselin, 1988). In the case of adoption of dairy technologies, the farmer's individual decision may be a function of spill over effects arising from neighbourhood effects like local weather and agro-climatic conditions, common terrain features and soil types, common sources of demand for milk and patterns of information diffusion. In case of spatial autocorrelation, the information content of the sample is lowered, rendering it less efficient than uncorrelated counterparts, so parameter estimates are inefficient, although asymptotically unbiased. Moreover, the omission of a spatially correlated and important variable may result in biased estimates (Anselin in this issue).

The application of spatial econometrics to household data is relatively uncommon, potentially due to the unavailability of geo-referenced household data. Methods to test and control for spatial effects have been mainly developed for the linear regression case. Spatial econometrics for limited dependent variables is a developing field of research and the methods developed to incorporate spatial dependence in these models are not of general applicability (Anselin and Florax, 1995, Nelson and Geoghegan, this issue). As Bockstael (1996) indicates, no satisfactory methods are available for addressing spatial autocorrelation in logit models. Although a number of authors have incorporated spatial effects in probit models (for example Case, 1991, and Case, 1991), the estimation techniques has not yet been included in spatial econometrics packages (SpaceStat, 2002 Website; S+ manual). If autocorrelation arises because some spatial process is not taken into account in the model, one way to control for spatial dependence is to include variables that account for interactions among farmers (Cressie, 1991; Odland, 1988). In this analysis, GIS-derived distances have been introduced to take into account farmers' market access. Calculated at the farmer-level, these variables potentially control for the occurrence of spatial autocorrelation by capturing the interactions between neighbours and potentially controlling for neighbours' influence on adoption (Case, 1991). Distances between farmers and a common point (i.e. urban centre) can be considered as an indirect measure of potential spatial relationships between farmers. It is thus postulated that the distances variables control for the existence of spatial autocorrelation. There are two arguments to support this hypothesis. First, using a method similar to that of Besag (1974) reported in Nelson and Geoghegan in this issue, the occurrence of spatial autocorrelation was tested by performing a random sampling on the original set of observations with the rule that the sampled farmers do not live in the same area. By sampling non-neighbours, it is reasonable to assume that the observations are spatially independent. Results from 500 iterations of sub-samples of one observation taken randomly from each sub-location are relatively consistent with those using the whole set of observations. While the results based on sub-samples may be biased because the data spatial structure is destroyed, the consistency between the sub-sample results and the overall results seems to indicate that spatial autocorrelation does not affect the results significantly. Secondly, in another analysis the authors applied spatial econometrics techniques on the same dataset to analyse milk price formation (Staal et al., 2000) since in the case of linear regressions used in that case, these techniques are available. Neighbours were defined as those farmers within a radius of 5 km. To ensure that the results do not depend on a specific distribution of the residuals, three regressions were conducted using different covariance structures (conditional spatial autoregression, simultaneous spatial autoregression and moving average). Spatial autocorrelation was tested using the Moran's I statistics. Results showed that residuals do exhibit spatial autocorrelation when only survey variables are introduced in the analysis but do not when both survey and GIS-derived variables, including distances, are included in the analysis, as in the case of the analysis presented in this paper. These two arguments thus suggest a low likelihood of significant inefficiency in the estimation results presented here, at least for the examples that include GIS-derived variables. The comparison estimates using only survey-based data may indeed suffer from this problem.

8.4. Model specification

Three technological choices are analysed. The first is the decision by a farmer to keep improved dairy cattle, defined as some level of cross between local Zebu and exotic European dairy stock, or pure exotic animals. These animals can require significant management expertise and resources. Their large size, susceptibility to local animal diseases, and need for specialised reproductive services mean that small farmers must invest considerable time and resources in producing or obtaining feed of the quantity and quality needed, treating or otherwise protecting their stock from disease, and obtaining artificial insemination or suitable bull service. The dependent variable is expressed in binary form, with 1 for the presence on farm of any improved dairy cattle, 0 otherwise.

The second technology choice considered is whether or not to grow specialised planted fodder to feed these animals. The fodder of choice in highland Kenya is Napier grass, which is easily established from locally-available cuttings, and yields high biomass. Its cultivation often requires that land be diverted from food or cash crops. If however, milk market opportunities offer positive incentives compared to the alternative crops, Napier may be grown even when land is scarce, so that most food may need to be purchased. Indeed, Napier cultivation yields more fodder per land unit than is available through grazed pasture (Kariuki, 1998), so that smaller land holdings are expected to increase likelihood of Napier cultivation. If farmers choose to grow Napier instead of pasture, animals are generally stall-fed, sometimes exclusively (zero-grazed), with the Napier then cut and brought to them. The dependent variable is expressed in binary form, with 1 for the cultivation on farm of Napier grass, 0 otherwise. The analysis is conducted only for dairy farms.

Finally, a third technology choice is considered, that of the supplementary feeding of purchased concentrate feed to dairy cows, generally only those lactating. These concentrates may be either commercial dairy meal, or grain milling by products, such as maize bran. Again, the derived demand for this input will depend on relative incentives for producing more milk. The improved dairy cattle are bred to respond to higher concentrate use by significantly higher milk production, if an adequate base diet of fodder is supplied simultaneously. However, credit and risks associated with it are important constraints, and research in Kenya has shown that even though higher concentrate use consistently raises average returns, risk-adjusted returns may often be lower when farmers' levels of risk aversion are incorporated (Kaguongo et al., 1997). Use of concentrates, which requires cash expenditure, will thus depend on expected milk prices and their variability, as well as production risks. Greater fodder